

Reconstructing Proto-Indo-European

David Langslow

Reconstruction plays an important part in many academic subjects, including every branch of classics. The business of every student of the classical world includes – or should include! – at least an element of reconstruction, aiming to answer either What was it like then? and What did it mean then? or What does it mean now? Look up ‘reconstruction’ in the *OED*, and you find a perhaps surprising amount of space given to the use of the word with reference to prehistoric languages. And it is here that David Langslow focuses in this article, with the oldest reconstructed object to which classicists contribute and of which they need to be aware, the reconstructed language Proto-Indo-European.

What is ‘Proto-Indo-European’?

The existence of Proto-Indo-European (PIE) – the prehistoric ancestor of Latin, Greek, English, and many other languages of India, Europe, and points in between – was postulated well over 200 years ago. The sentences most often quoted as the first claim of the genetic relatedness of the Indo-European (IE) languages are from a lecture given in 1786 by Sir William Jones, a supreme court judge and amateur scholar, to the Asiatic Society in Kolkata (Calcutta), then the capital of the British Raj. He is talking about Sanskrit, the classical language, the ‘Latin’, of India, which westerners were learning in growing

numbers as the cultural treasures of India were opened up to the European powers, and he says this about what strikes you when you learn Sanskrit after learning Greek and Latin at school:

The Sanscrit language, whatever may be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed that no philologist could

examine them all three without believing them to have sprung from some common source, which perhaps no longer exists.

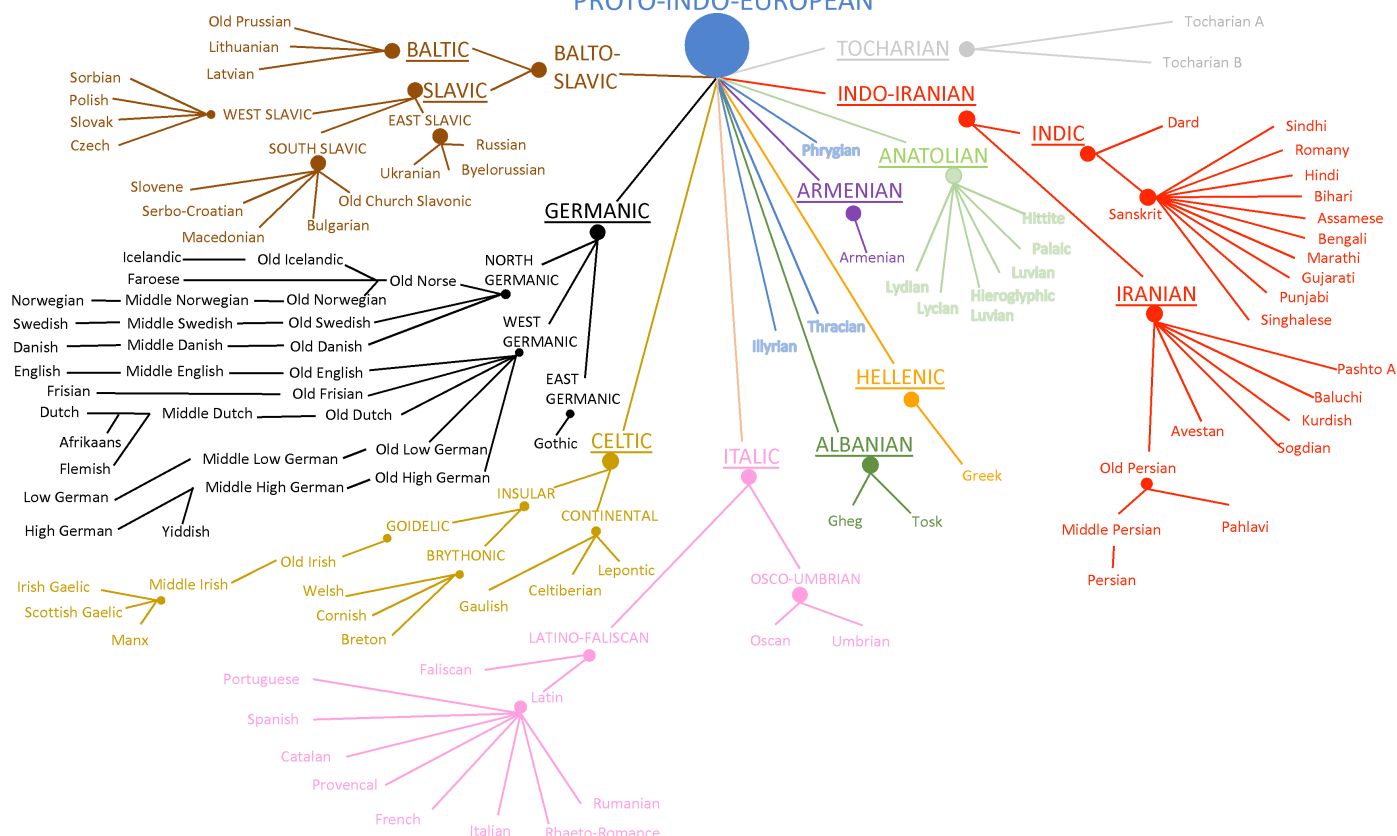
It is interesting that Jones talks not of vocabulary but of grammar. He could so easily have drawn attention to Sanskrit words that look awfully like Latin or Greek (or English!), words such as *mātar-* ‘mother’, *pitar-* ‘father’, *dant-* ‘tooth’, *pad-* ‘foot’, *trayas* ‘three’, *navas* ‘new’, and so on. But he doesn’t, he talks instead of ‘affinity in the roots of verbs and in the forms of grammar’. What sort of thing does he have in mind? One tiny but telling instance is this:

	Sanskrit	Latin	Gothic
‘is’	as-ti	es-t	is-t
‘are’	s-anti	s-unt	s-ind

‘Is’ and ‘are’ in Sanskrit, Latin, and Gothic (an early Germanic language, a distant cousin of English).

In the forms for ‘is’ and ‘are’ in three separate languages we note the remarkable facts that all three languages agree in four details: (1) they show an *s* in the first part, the root, of both forms, (2) at the start of the word in the ‘are’ form but preceded by a short vowel in the ‘is’ form; and (3) they use endings which contain a *t* in the ‘is’ form, and (4) a sequence of short vowel + *n* + *t* or *d* in the ‘are’ form. Jones’s instinctive belief in ‘some common source’ has been recently borne out by calculations of statistical probability applied to thousands of known languages. It emerges the probability of the agreements over ‘is’ and ‘are’ arising independently by chance in more than one language is minute, or, in other words, any languages that have them are almost certainly showing retentions from a unique common ancestor-language and are, therefore, related, sister-languages. After this first step, the hypothesis of relatedness, comes the business of reconstruction: of deciding which languages belong to the family and of how they may be grouped in sub-families; and of postulating what the ancestor language itself actually looked like.

PROTO-INDO-EUROPEAN



‘Family tree’ of Indo-European languages.

Attested languages ancient and modern (in lower case in the family tree above) known to descend from PIE, much as the Romance languages descend from Latin, are grouped in twelve branches (underlined in the family tree). Of course, we have not a single document in PIE; everything that we say about the language and its speakers is a part of a grand and elaborate hypothesis. This hypothesis rests on subjecting the sounds, forms, words, and sentences of the daughter-languages to various means of reconstruction.

The ‘Comparative Method’

It is on the comparative method that the reconstruction of any ancestor or ‘proto-language’ fundamentally rests. The method was developed first for PIE towards the end of the nineteenth century, and ever since then it has been used in studies of language families worldwide for establishing relatedness between languages and reconstructing their common ancestors. The method is based on one assumption and on one remarkable fact.

The assumption is that, apart from onomatopoeic words (like *woof*, *miaow*, *splash*, *bang*), in all natural languages there is an arbitrary relation between sound and meaning; there is nothing in the meaning *THREE*, for instance, that calls for, determines, or is represented by the sound or form of English *three*, French *trois* or German *drei*. It follows that the use by different languages of the same or similar forms for the same or similar meanings is the result either of chance, or of borrowing, or of common inheritance. Chance agreements, or near-agreements, of form and meaning across languages do occur but are extremely rare.

The comparative method takes for its data agreements of a very special sort, what amount to agreements between languages in rhyming and alliterating words. This has to do with the remarkable fact that, unlike other types of language change, changes in the pronunciation of vowels and consonants in a given language tend to be absolutely regular. If, say, the Old English sound /ā/ (as in *ah*) changes to /əu/ (as in *owe*) in Modern English in one word

in which it occurs, then, other things being equal, the sound /ā/ will change to /əu/ in every word in which it occurs. And if the Old English sound cluster /kn/ (as in *banknote*) changes to /n/ at the start of one Modern English word in which it occurs, then /kn/ will change to /n/ at the start of every word where it occurs. In other words, we might say, ‘once a rhyme, always a rhyme; once an alliteration, always an alliteration’. Compare the Old English and the Modern English rhymes and alliterations below – and ignore modern spelling, especially of *kn*!; focus on the pronunciation.

Old English	āc	fā	hāl	hlāf	gāt	ān-	stān
Modern English	oak	foe	whole	loaf	goat	on-ly	stone
Old English	cneow	cnif	cnihht	cnotta	cnawan	cnedan	cnyttan
Modern English	knee	knife	knight	knot	know	knead	knit

Two examples of regular sound-change: Old English ‘ah’ becomes Modern English ‘oh’, and Old English /kn/ becomes Modern English /n/!

So what?!

Very interesting, you may be thinking, but where does this get us? How does ‘once an alliteration, always an alliteration’ serve as a basis for reconstructing a prehistoric language? Here’s how. I invite you to consider the words and especially the highlighted consonants in the table over the page. The words arranged in columns are from English, Latin, Greek, and Sanskrit. The words

in each row are synonyms, or very nearly so: the meaning is virtually the same across each row.

English	Latin	Greek	Sanskrit
<u>th</u> ree	trēs	treîs	tráyas
f <u>ath</u> er	p <u>a</u> ter	patér	pitár
f <u>oo</u> t	ped-	pod-	pád-
t <u>en</u>	decem	déka	dása
t <u>oo</u> th	dent-	odont-	dant-
b <u>ro</u> ther	frāter	phrātér	bhrātar
b <u>ear</u>	fer-	phér-	bhar-

Repeating alliterations between very different languages, attested hundreds and thousands of years apart, and hundreds and thousands of miles apart!

Notice what the underlinings indicate, that what we are actually comparing across languages are not whole words of similar form and meaning but individual sounds within them.

A remarkable observation

Sets of data like that above allow remarkable statements of the following kind: ‘In these words of similar form and meaning, wherever I find the English *th* sound, I find Latin *t*, Greek *t*, Sanskrit *t* (look at the words for three, father, tooth, brother, with this pattern in bold); wherever I find English *t*, I find Latin *d*, Greek *d*, Sanskrit *d* (consider the words for foot, ten, tooth); and so on.’ More formally we can draw up correspondence-sets, like those below, which capture all of the consonant agreements in the examples above.

English	Latin	Greek	Sanskrit	
/θ/	/t/	/t/	/t/	(< PIE *t)
/f/	/p/	/p/	/p/	(< PIE *p)
/r/	/r/	/r/	/r/	(< PIE *r)
/t/	/d/	/d/	/d/	(< PIE *d)
/h/	/k/	/k/	/ś/	(< PIE *k)
/b/	/f/	/ph/	/bh/	(< PIE *bh)

Regular sound-correspondences between the four languages.

The only possible explanation

Each of these sound-correspondences is repeated (you have to take my word for this: trust me, I’m a doctor!) in dozens of sets of formally similar (near-)synonyms in languages belonging to all twelve branches of the IE family; the ‘same’ words alliterate in Irish and Hindi, in Swedish and Persian, in Latvian and Welsh. Now, a quick calculation will show that the probability of this state of affairs arising by chance is infinitesimal; borrow-

ing is out of the question as an explanation: languages can borrow virtually anything, but they just don’t borrow rhymes and alliterations. The only possible explanation is that these sounds have indeed, in Jones’s phrase, ‘some common source’ from which they have developed regularly but differently in each language.

Reconstruction

OK, but we still haven’t reconstructed anything. All we’ve established is that agreements in alliterations between languages reflect a single prehistoric consonant changing regularly but differently in each language. Right! So the next step is to consider which single prehistoric consonant we are to reconstruct as the common source of, say, Eng *th* : L *t* : Gk *t* : Skt *t*, or (much trickier) of Eng *b* : L *f* : Gk *ph* : Skt *bh*. How did the prehistoric speaker pronounce the first consonant of the word for three, the middle consonant of the word for father, and the last consonant of the word for tooth? – as *t*, as *th*, or as something else? In principle, we could call it *x*, or ‘sound 43’ (and some sounds are referred to by number, as we shall see); in practice, what we reconstruct is guided sometimes by a ‘majority vote’ among the witness languages, always by the plausibility of the entailed sound-changes. From the *th* : *t* : *t* correspondence-set above, we reconstruct the ancestral (PIE) sound **t* (as you see in the right-hand column of the table on the left), partly because most languages have *t*, and partly because the change *t* → *th* is common, while the converse (*th* → *t*) is not. (By the way, the asterisk in **t* and throughout that table marks unattested, hypothetical reconstructed forms.)

That’s it, really. The rest is down to patient repetition, hard work, and a lot of arguing, first over isolating all the recurring alliterations, rhymes, and sound-correspondences between IE languages, and then over which single PIE sound to reconstruct as the common source of each. Yes, individual sounds in isolation are meaningless, but in combination they acquire meaning, and Indo-Europeanists can agree on quite a detailed account of PIE grammar and vocabulary. This means that we can reconstruct the PIE roots of the words used in the table above: **treyes* ‘three’ m. nom. pl.; **pəter-* ‘father’; **pēd-* ‘foot’; **dek’m* ‘ten’; **dēnt-* ‘tooth’; **bhrāter-* ‘brother’; **bher-* ‘carry’.

In 1868 a German scholar named August Schleicher answered the challenge, ‘If PIE was a real language, what did it look like?’ with this fable:

**G^werei owis, k^wesyo wlhnā ne ēst, ekwons espekēt, oinom ghe g^wrum woghom wegfontm, oinomk^we megam bhorom, oinomk^we ghmenm*

ōku bherontm. Owis nu ekwobhos ewewk^wet: ‘Kēr aghnutoi moi ekwons agontm nerm widntei.’ Ekwōs tu eweuk^wont: ‘Kludhi, owei, kēr ghe aghnutoi nsmei widntbhos: nēr, potis, owiōm r wlhnām sebhei g^whermom westrom k^wrneuti. Neghi owiōm wlhnā esti.’ Tod kekluwōs owis agrom ebhuget.

(‘Now, a sheep, which had no wool, saw some horses, one pulling a heavy wagon, one a big load, and one carrying quickly a man. The sheep spoke to the horses: “The heart suffers in me as I see the man driving the horses.” And the horses spoke: “Listen, sheep, the heart suffers in us who know: the man, the master, makes sheep’s wool into warm clothes for himself. And the sheep have no wool.” On hearing this, the sheep fled into the field.’)

If you know some Latin or Greek, you can have fun spotting words you know which derive from the PIE forms in the story! Two things may have caught your eye in this story and in the PIE words above. First, **l*, **m*, and **n* stand often between consonants (**wlhnā*, **widntei*) or at the end of the word after a consonant (**bherontm*, **agontm*), where they must be pronounced almost as vowels (compare the *m* in English words ending in *-ism*, or the *l* in *bottle*). Second, what about the **ə*₂ in the word for father, related to the **h* between consonants in the word for wool (**wlhnā*) in the fable? Remember my warning above that some PIE sounds are given numbers! Well, most of what we’ve seen so far of PIE is from a late stage of the language, shortly before it all broke up and people went off in different directions at some point in the third millennium B.C. or earlier. These odd-looking sounds (the vowels **l*, **m*, **n*, **r*, and even more **h* and **ə*₂) are from an older phase of PIE. – But, hang on! How on earth can you say anything about an older phase of an already prehistoric reconstructed language?! – Well, it was the astounding achievement of a Swiss linguist in his twenties to invent (in the 1870s) a new method of reconstruction, internal reconstruction (i.e. reconstruction within a single language), and to apply it – for its very first outing – to late PIE so as to reveal an earlier stage of this, already prehistoric, language.

To be continued...

David Langslow is a doctor (of Philosophy); indeed, a Professor – of Classics, at Manchester, where he specializes in ancient medical texts and language, and descriptive and social linguistics. He is one of the directors of the JACT Greek summer school.